

Tsallis distribution as a standard maximum entropy solution with ‘tail’ constraint

J.-F. Bercher¹

*Université Paris-Est, LabInfo-IGM, 5 bd Descartes, 77454 Marne la Vallée Cedex 2, France
and Laboratoire des Signaux et Systèmes, CNRS-Univ Paris Sud-Supelec, 91192 Gif-sur-Yvette Cedex, France*

Abstract

We show that Tsallis’ distributions can be derived from the standard (Shannon) maximum entropy setting, by incorporating a constraint on the divergence between the distribution and another distribution imagined as its tail. In this setting, we find an underlying entropy which is the Rényi entropy. Furthermore, escort distributions and generalized means appear as a direct consequence of the construction. Finally, the “maximum entropy tail distribution” is identified as a Generalized Pareto Distribution.

Key words:

Maximum entropy principle, Rényi entropy, nonextensive statistics

PACS: 89.70.cf, 02.50.-r, 05.90.+m

1. Introduction

The maximizers of a special entropy, the Tsallis entropy [1], with suitable constraints, are often called Tsallis distributions. It is worth mentioning that the maximization of any monotoneous transform of Tsallis entropy, with the same constraints, leads to the same maximizers. This is in particular the case of Rényi entropy. In applied fields, Tsallis distributions (q -distributions) have encountered a large success because of their remarkable agreement with experimental data, see for instance [2,3] and references therein. These distributions are of very high interest in many physical systems, since they can exhibit heavy-tails, and model power-law phenomena. Indeed, power-laws are especially interesting since they appear widely in physics, biology, economy, and many other fields [4]. Tsallis distributions are similar to Generalized Pareto Distributions, which also have a high interest in other fields, namely reliability theory [5], climatology [6], radar imaging [7] or actuarial sciences [8]. Hence, a remarkable feature of the maximum Tsallis entropy construction is its ability to exhibit heavy tailed distributions. Furthermore, the essence of Pickands’ extreme values theorem [9] is that the distribution of excesses over a threshold converges, under wide conditions, to a q -exponential [10]. Following the idea that an interest of Tsallis distributions is in fact *a tale of tails*, in the words of [11], we suggest that this kind of distributions can also be obtained from the familiar (Shannon) maximum entropy setting, by the introduction of an appropriate constraint. In this setting, we show that Rényi entropy appears naturally, and also obtain a natural interpretation of nonextensive escort distributions, ‘generalized means’ and entropic index.

Email address: `jf.bercher@esiee.fr` (J.-F. Bercher).

¹ On sabbatical leave from ESIEE-Paris, France.

2. Maximum entropy with ‘tail’ constraint

Jaynes’ maximum entropy principle [12,13] suggests that the least biased probability distribution that describes a partially-known system is the probability distribution with maximum entropy compatible with all the available prior information. The Kullback-Leibler information divergence $D(P||Q)$ measures the divergence of a distribution P to another distribution Q .

2.1. Problem setting

We call ‘tail distribution’ the probability density function (pdf) of the excesses X_u of a variate X over a threshold u . It is not only the tail part of the distribution, but is shifted by u toward the origin, and, as a pdf, is normalized to 1. Let now P and Q be two probability density functions. As a guideline, we find useful to imagine P as the tail distribution of Q : in such a case, these distributions are closely related. Our idea is to account more generally for an existing relationship between two distributions. Therefore, we propose to parametrize the relation between a candidate distribution and its ‘parent’ by the value θ of the Kullback-Leibler divergence between them: $D(P||Q) = \theta$. This define a set of possible distributions, and the solution is selected in this set, according to the maximum entropy principle, as the distribution P with maximum entropy. This writes as follows:

$$\begin{cases} \max_P H(P) = - \int P(x) \log P(x) dx \\ \text{s.t. } D(P||Q) = \int P(x) \log \frac{P(x)}{Q(x)} dx = \theta. \end{cases} \quad (1)$$

The definition of $D(P||Q)$ requires that P is absolutely continuous with respect to Q . It is understood, as usual, that $0 \log 0 = 0 \log 0/a = 0 \log 0/0 = 0$ and that $a \log(a/0) = +\infty$ if $a > 0$.

An alternative formulation, which leads to the same solution, could be to look for the distribution with minimum divergence to Q , in the set of all distributions with a given entropy:

$$\begin{cases} \min_P D(P||Q) \\ \text{s.t. } H(P) = \theta'. \end{cases} \quad (2)$$

It is clear that the optimum distribution arising from this procedure will not be the ‘exact’ tail distribution, since the distribution of the excesses $X_u = X - u | X > u$, which reads $X_u = X - u$ if $X > u$ (conditioned variable), has the pdf $P_u(x) = Q(x + u)/\bar{Q}(u)$, for $x \geq 0$, and where $\bar{Q}(u) = \int_u^{+\infty} Q(x) dx$ is the so-called survival function. It is rather the maximum entropy variant, when the ‘tail distribution’ P is constrained to be at given divergence θ to the parent distribution Q . In fact, we do not need to rely on tails in this construction, but we simply introduce a constraint on the Kullback-Leibler divergence to a ‘parent’ distribution in order to account for a general relationship between these distributions.

2.2. The maximum entropy solution and first consequences

The solution of (1) is easily derived using standard maximum entropy results, e.g. [14]. In this case, we obtain

$$P(x) = \frac{1}{Z(\lambda)} e^{\lambda \log \frac{P(x)}{Q(x)}} \quad (3)$$

with λ the Lagrange parameter associated to the constraint $D(P||Q) = \theta$, and $Z(\lambda)$ the partition function. It can also be reduced to

$$P(x) = \frac{1}{Z(\lambda)^{\frac{1}{1-\lambda}}} Q(x)^{\frac{\lambda}{1-\lambda}} \quad (4)$$

with

$$\log Z(\lambda) = (1 - \lambda) \log \int Q(x)^{\frac{\lambda}{1-\lambda}} dx. \quad (5)$$

As usual in the maximum entropy setting, the constraint and partition function are linked by

$$\frac{d \log Z(\lambda)}{d\lambda} = \frac{d}{d\lambda} \log \int e^{\lambda \log \frac{P(x)}{Q(x)}} dx = \theta. \quad (6)$$

Let us now denote

$$q = \frac{\lambda}{\lambda - 1}. \quad (7)$$

Then, it clearly appears on the one hand that the ‘maximum entropy tail’ (4), which can be rewritten as

$$P(x) = \frac{Q(x)^q}{\int Q(x)^q dx}, \quad (8)$$

is the *escort distribution* of nonextensive thermostatics. On the other hand, the log-partition function (5) becomes

$$\log Z(q) = \frac{1}{1-q} \log \int Q(x)^q dx \quad (9)$$

where the right-hand side has exactly the form of the *Rényi entropy* of distribution Q . It shall be mentioned that the Rényi entropy has no definite concavity for $q > 1$ and is not Lesche stable [15,16]. Its use in statistical physics has been discussed, e.g. [17]. In our context, it appears as a by-product of our original problem (1) which involves the maximization of the standard Boltzmann-Shannon entropy.

For the optimum distribution, we can also observe that the maximum (Shannon) entropy reduces to

$$H(P) = - \int P(x) \log P(x) dx \quad (10)$$

$$= - \int P(x) \left(\lambda \log \frac{P(x)}{Q(x)} - \log Z(\lambda) \right) dx \quad (11)$$

$$= -\lambda \int P(x) \log \frac{P(x)}{Q(x)} dx + \log Z(\lambda) \quad (12)$$

$$= -\lambda\theta + \log Z(\lambda) \quad (13)$$

where the last relation is obtained using the definition of θ in (1).

Therefore, we obtain that the maximum entropy problem (1) has for optimum value the Rényi entropy (9) with index q , minus a linear function of the constraint.

2.3. Solution with an additional observation constraint

Suppose now that the original problem is completed by an additional constraint. Indeed, the definition of equilibrium distributions usually need to take into account observation constraints. For instance, one often has to account for an observable defined as a mean value under distribution P . This is very classical in maximum entropy approaches. The constraint writes

$$m = E_P [X]. \quad (14)$$

Since P is the escort distribution (8) of Q , this mean constraint is also the *generalized mean* constraint of nonextensive thermostatics:

$$m = \int xP(x)dx = \frac{1}{\int Q(x)^q dx} \int xQ(x)^q dx. \quad (15)$$

When P is thought as the ‘tail’, m is the well-known *mean residual lifetime* (in reliability theory), or *expected future lifetime* (in survival analysis).

Because of the additional constraint, determination of the ‘maximum entropy tail’ distribution P amounts to further maximize the entropy in (13) subject to that constraint. Using (9), we obtain

$$\begin{cases} \max_Q \frac{1}{1-q} \log \int Q(x)^q dx \\ \text{s.t. } E_P [X] = m, \end{cases} \quad (16)$$

which leads to

$$Q(x) \propto (1 - \beta(1 - q)(x - m))^{\frac{1}{1-q}} \quad (17)$$

for the value of the parameter β such that $E_P[X] = m$, with P given by (8). As far as the latter is concerned, its expression is simply

$$P(x) \propto (1 - \beta(1 - q)(x - m))^{\frac{q}{1-q}}. \quad (18)$$

This relation can also be rearranged as

$$P(x) \propto (1 + \beta(q - 1)(x - m))^{-\frac{1}{q-1}-1}, \quad (19)$$

which is exactly in the form of the Generalized Pareto Distribution [18], with shape factor $(q - 1)$ and scale factor β . Using (7), the exponent in (19) reduces to $-\lambda$, so that the distribution asymptotically behaves as a power-law with exponent $-\lambda$.

Observe that with $\theta = 0$ in (1), we readily have $P = Q$ and $q = 1$, and P in (19) reduces to the classical Boltzmann-Gibbs canonical distribution.

As a final comment, let us note that the maximum entropy tail distribution is *stable* with respect to thresholding, in the sense that the result remains in the same family, with entropic index q but a different scale parameter. If u is the threshold, then the pdf of excesses over u , with P the parent distribution, is

$$P_u(x) \propto P(x + u) \propto (1 + \beta(q - 1)(x + u - m))^{-\frac{1}{q-1}-1} \propto (1 + \beta'(q - 1)(x - m))^{-\frac{1}{q-1}-1}, \quad (20)$$

where the last term, with $\beta' = \beta/(1 + \beta u)$, is obtained by factoring $(1 + \beta u)$. This highlights the particular status of the ‘maximum entropy tail distribution’ as a tail distribution.

3. Conclusion

In this Letter, we followed two ideas. First, that an important feature of Tsallis’ distributions is their ability to model the tail of distributions, particularly of those with heavy tails. The second idea is that these distributions should appear in a standard maximum entropy setting. This led us to the introduction of a constraint on the divergence between two distributions, one of them being imagined as the tail distribution. This constraint accounts for an existing general relationship between two distributions.

We showed that within this construction, the escort distributions and generalized means of nonextensive statistics appear very naturally. We also obtained that the original maximum (Shannon) entropy reduces to a maximum Rényi entropy; or equivalently to the maximization of Tsallis entropy. As far as the entropic index q is concerned, it is simply associated to the value of the divergence between the distribution Q and its escort distribution P . Finally, the ‘maximum entropy tail’ distributions, in the sense adopted in this Letter, are found to be Generalized Pareto Distributions, which have proved very useful for modeling heavy-tailed distributions in many applied problems. We believe that this construction can be useful to workers in the field.

References

- [1] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, *Journal of Statistical Physics* 52 (1988) 479–487.
- [2] C. Beck, Generalized statistical mechanics of cosmic rays, *Physica A* 331 (2004) 173–181.
- [3] C. Tsallis, Entropic nonextensivity: a possible measure of complexity, *Chaos, Solitons, & Fractals* 13 (2002) 371–391.
- [4] M. E. J. Newman, Power laws, Pareto distributions and Zipf’s law, *Contemporary Physics* 46 (2005) 323–351.
- [5] M. Asadi, I. Bayramoglu, The mean residual life function of a k-out-of-n structure at the system level, *IEEE Transactions on Reliability* 55 (2006) 314–318.
- [6] M. A. J. V. Montfort, J. V. Witter, Generalized Pareto distribution applied to rainfall depths, *Hydrological Sciences Journal* 31 (1986) 151–162.
- [7] B. LaCour, Statistical characterization of active sonar reverberation using extreme value theory, *IEEE Journal of Oceanic Engineering* 29 (2004) 310–316.
- [8] A. C. Cebrian, M. Denuit, P. Lambert, Generalized Pareto fit to the society of actuaries’ large claims database, *North American Actuarial Journal* 7 (2003) 18–36.
- [9] J. Pickands, Statistical inference using extreme order statistics, *The Annals of Statistics* 3 (1975) 119–131.
- [10] J.-F. Bercher, C. Vignat, A new look at q -exponential distributions via excess statistics, *Physica A* (2008) doi:10.1016/j.physa.2008.05.038, in press.

- [11] Montroll, Shlesinger, Maximum entropy formalism, fractals, scaling phenomena, and $1/f$ noise: A tale of tails, *Journal of Statistical Physics* 32 (1983) 209–230.
- [12] E. T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.* 108 (1957) 171.
- [13] E. T. Jaynes, On the rationale of maximum entropy methods, *Proc. IEEE* 70 (1982) 939–952.
- [14] I. Csiszár, Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems, *Annals of Statistics* 19 (1991) 2032–2066.
- [15] B. Lesche, Instabilities of Rényi entropies, *Journal of Statistical Physics* 27 (1982) 419–422.
- [16] S. Abe, Stability of Tsallis entropy and instabilities of rényi and normalized Tsallis entropies: A basis for q -exponential distributions, *Physical Review E* 66 (2002) 46134.
- [17] P. Jizba, T. Arimitsu, Observability of Rényi's entropy, *Physical review. E* 69 (2004) 026128, pMID: 14995541.
- [18] N. L. Johnson, S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions*, 2nd Edition, Wiley-Interscience, 1994.